Forecast Visualizations for Terrorist Events

*Jason Goffeney** Naval Research Lab / ITT Corporation
Greg S. Schmidt† Naval Research Lab
Jason Dalton‡ Spatial Data Analytics (SPADAC)
Jim D'Archangelo§ U.S. Naval Academy
Ruth Willis¶ Naval Research Lab

ABSTRACT

We are developing techniques to forecast terrorist events and effective ways to present these forecasts to intelligence analysts. Forecasts come from analyzing historical event data and geographical information. We explore feature reduction techniques to make the computations closer to real-time and techniques for representing the confidence (or uncertainty) of the data.

CR Categories: G.3 [Mathematics of Computing]: Probability and Statistics—Probabilistic Algorithms; H.5.0 [Information Systems]: Information Interfaces and Presentation—General;

Keywords: forecast, uncertainty, prediction, PDF, visualization

1 INTRODUCTION

Having the ability to forecast terrorist events is of utmost importance to intelligence analysts and military planners performing counter-measures for the global war on terror. We are currently developing techniques to forecast the likeliest locations a terrorist would target. We are extending earlier work [1] that utilizes historical event and geographic information system (GIS) information data to generate geospatial likelihood functions indicating where an attack may occur next. Part of our effort is focused on the computationally-intensive problem of reducing the search space produced by the large amounts of GIS and event data. We also explore how to represent the confidence of the data by assessing and characterizing the types of uncertainty and developing effective presentation approaches. We consider the impact of having error in the historical event and feature data, choice of feature reduction method, and choice of likelihood function.

We describe briefly our progress in developing techniques for feature reduction, event forecasts, and associated display techniques. We also highlight our current plans to include confidence (uncertainty) information into the forecast visualizations.

2 FEATURE REDUCTION

One of the challenges of working with comprehensive GIS layers is the vast number of features available for consideration. Our data ranges from just a few embassy locations to thousands of street junctions. Because the events, usually bombings, are scattered across the area of interest, it is not immediately apparent which features are significant. The benefit of feature reduction is not just to eliminate extraneous and possibly misleading pieces of information, but to also improve computational memory and time requirements.

The simplest methods are to limit the number of features to consider based on certain metrics (such as a maximum distance from the event) or constraining each feature to lie within a regional bounding box. Both methods assume terrorists prefer certain spatial features (consciously or not), such as buildings or streets near the target location. The initial results are promising, although we do not have a clear understanding of the distance at which features remain viable for target selection.

Several numerical approaches are being reviewed such as principal components, clustering, and factor analysis. Currently we are working with the Gini index [2]. The purpose of this method is to provide a ranking for each feature based on inter-event distances. Each event is represented as a vector of spatial distances from its location to the location of each feature in the layers. If d_{ij} is the numeric distance between events i and j for the same vector element then its similarity $s_{ij}$is calculated as:

$$s_{ij} = \frac{1}{1+\alpha d_{ij}}$$

where $\alpha = 1/\bar{d}$ and $\bar{d}$ is the average numeric inter-event distance. Note that spatial distance refers to the distance between an event location, while numeric distance refers to the "distance" between two vector elements as a distance between distances.

The Gini index between two events is defined as

$$g_{ij} = 4s_{ij}(1-s_{ij}).$$

For the entire set of events the averaged Gini index

$$I_g = \frac{2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} g_{ij}}{n(n-1)}.$$

is a suitable measure of cohesiveness. A lower value of I_g is considered to denote similarity within the feature space. A further extension of the method adjusts this value based on its disparity from the background distribution for this feature. For example, if every event occurs within 50 m of an ATM, but every ATM location in the area is within 50 m, then the distance to the ATMs is not a very useful measurement. The number of features can be reduced by establishing a cutoff threshold for the Gini index. An example of this reduction is provided in Figure 1.

3 EVENT FORECASTING

The problem of determining spatial preferences has been successfully applied to urban settings to find potential crime hot spots by looking at factors such as economics, populations, proximity of key building types, and past criminal histories. Brown et al. [1] applied the same technique to look at terrorist event preferences. We are using this work as a roadmap for our efforts. Our goals are to develop forecast image overlays for regional maps of the target locations that predict the likeliest locations of terrorist events. The combined map and overlay will aid security operations in determining the best places to deploy security forces or sensing equipment.

We employ a few different methods to generate forecast overlays for the geographic region of interest. One of these methods is the Gaussian-based forecasting technique derived in [1]. The premise of the technique is that a suicide bomber is directed toward a certain location by a set of qualities such as geospatial features, demographic information, and recent political events. Focusing on the geospatial domain, we consider the intended target was associated with the features located within a small distance from the event location. Furthermore, we consider the distance between key features

*jason.goffeney@nrl.navy.mil
†gregory.schmidt@nrl.navy.mil
‡dalton@spadac.com
§jmd@usna.edu
¶ruth.willis@nrl.navy.mil

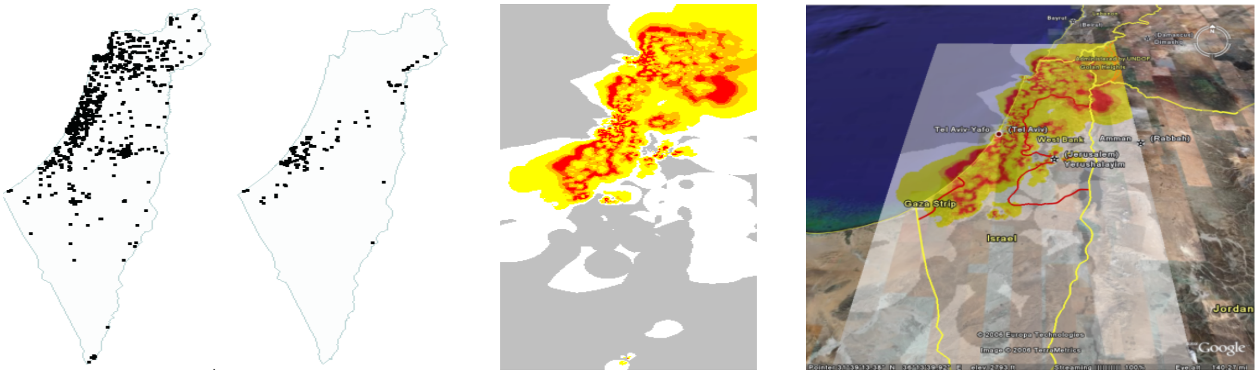

Figure 1: (Left) Before and after the Gini index feature reduction technique is applied to embassies and gas station features. (Right) Likelihood of terrorist attacks using (1) GIS information about locations of embassies and gas stations, and (2) historical terrorist events between 2001-2004. Forecast layer generated by our testbed, converted to KML, and loaded into Google Earth.

and the event location as the highest likelihood, and taper the likelihood values as the distances increase or decrease away. We model this effect using a Gaussian distribution centered at the distance between key features and the event. The probability density function (PDF) for a feature i for a given grid cell g is given by

$$f(D_{ig}) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{\frac{-(D_{ig}-D_{in})^2}{2\sigma_i^2}}$$

where D_{ig} is the distance from the feature to the grid cell, D_{in} is the distance from the feature to event location n, and N is the total number of events.

The joint density for the entire feature set is established by the product of each density result as

$$f(\rho_g) = c\prod_{i=1}^{I}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{\frac{-(D_{ig}-Din)^2}{2\sigma_i^2}}$$

where I is the total number of features and c is a constant.

Another method implemented and explored was a k-nearest neighbors approach: take the set of grid vectors for each feature, calculate k minimum distances to the event vectors, and keep the median. The reasoning is that if a terrorist has a vector of preference to a certain geospatial arrangement then any grid cell that is similar to it should be a likely candidate for a terrorist attack.

4 Confidence Modeling

One of the most important aspects of forecasting is having an estimate of the confidence in the supporting numerical values. In numerical weather prediction, there is always some value of certainty associated with the forecasts. One example is a prediction of 80% chance of rain, which implies that the numerical weather modeler(s) predicted that 8 in 10 times it would rain tomorrow. Having confidence (or uncertainty) associated with the terrorist event forecasts is very important. We have identified several sources of uncertainty that must be modeled for each event forecast. We feel the most important sources are: positional uncertainty associated with the GIS and event data, error associated with the feature reduction, and error in the choice of event prediction technique (i.e., error associated with generating the likelihood functions). For now, we are only beginning to model the positional error of the event locations, for which we briefly describe our approach and show a mock-up of the resulting visualization technique we plan to use.

Each historical event record contains the date, location, type of attack, organization who claimed responsibility, a description of what happened, and confidence of the recorded data. The confidence values for the location are rated from 1 to 5, with error values starting at ± 1 m and increasing by a power of 10 for each rank. The values represent the uncertainty about the exact location of a detonation as the analysts try to extract the information from news sources. This location uncertainty impacts the computation of the distances computed from each event to the nearby features (e.g., building, street intersections, subway stations, etc.). The distances become a range of values $D_{in} \pm E(r)$, where r is the rating index and $E(r)$ is the error value. Accounting for this variation, a range of PDFs result for each event location used in the computations.

We plan to start by using the distances associated with the maxima and median of the range of error (3 distances), producing $3n$ times as many PDFs. The first visualizations will likely use an interface slider to page through the resulting PDFs. The second visualization will aggregate the highest risk locations fusing them into one image. A third approach will use the median values to generate the main PDF, and then use elevation to show the error (or range of values) associated with the minima and maxima.

5 Results and Conclusions

We have developed a software testbed for the algorithms using Trolltech's Qt library (*www.trolltech.com*) combined with OpenGL. We also generate forecast layers in the KML syntax (*earth.google.com/kml*) and display them using Google Earth (example shown in Figure 1 (right)). The testbed supports GIS and historical event database formats: Microsoft Access and ESRI (*www.esri.com*).

To conclude, we have explored several techniques for performing feature reductions, developing forecasts, and proposed several techniques for incorporating confidence information into the visualizations of the forecasts. Our efforts are ongoing and include plans to explore other feature reduction methods (e.g., parallel algorithms), other likelihood functions (e.g., likelihood ratios involving one faction not being involved,

$$LR(x|f) = \frac{P(x|f)}{P(x|g)},$$

where g represents a faction known not to be involved with the event, f represents all the factions), and methods for representing confidence (or uncertainty). We are also beginning to explore the appropriateness of using Bayesian analysis with Gibbs Sampling as a tool in our approach.

References

[1] D. Brown, J. Dalton, and H. Hoyle. Spatial forecast methods for terrorist events in urban environments. In *Symposium on Intelligence and Security Informatics (ISI)*, volume 3073, pages 426–435, June 2004.

[2] H. Liu and D. Brown. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 19:603–622, 2003.